



## بخشی از ترجمه مقاله

عنوان فارسی مقاله :

استخراج مجموعه اقلام مکرر در یک جریان

عنوان انگلیسی مقاله :

Mining frequent itemsets in a stream



### توجه !

این فایل تنها قسمتی از ترجمه میباشد. برای تهیه مقاله ترجمه شده کامل با فرمت ورد (قابل ویرایش) همراه با نسخه انگلیسی مقاله، [اینجا](#) کلیک نمایید.



## بخشی از ترجمه مقاله

## 8. نتیجه گیری

## 8. Conclusion

The max-frequency measure is apt for mining frequent itemsets in a datastream. It truthfully captures the frequency of itemsets, as is shown by Fig. 10. Frequent itemsets can be mined efficiently, both in terms of time and space. The efficiency of the algorithm can be derived theoretically by linking the number of borders to the concept of Farey streams, from number theory. From this connection we can deduce that, even in the worst case, the number of borders increases slower than the length of the stream. Experiments on datasets, in which itemsets occur in varying patterns, show that the number of borders in a summary, in practice, remains relatively stable over time.

We showed that no edge can be gained from exploiting the subset relation between itemsets. Simply storing item-summaries results in space issues. Sharing borders among summaries linked by the subset relation on itemsets is impossible because at any given point a position might be a border for itemsets  $A_1$  and  $A_3$  but not for itemset  $A_2$ , where we have that  $A_1 \subset A_2 \subset A_3$ . As a result, further optimizations of the algorithm will need to be based on non-trivial insights.

From the experiments with our software prototype on both synthetic and real-world datasets, we draw three conclusions. Firstly, on average very few borders are needed to derive the max-frequency of all itemsets in a stream. Secondly, streams that are sparse & bursty over the whole course of the stream benefit from the lazy handling checkscheduler. Thirdly, even stream with large transactions can be mined efficiently if suitable values for the minimal window length and minimal frequency threshold have been set.

اندازه‌گیری ماکسیمم فرکانس برای کاوش مجموعه آیت‌های مکرر در یک جریان داده‌ای، مناسب است. آن به گونه‌ای صادقانه فرکانس مجموعه آیت‌ها را در بند می‌کشد همانطور که در شکل ۱۰ نشان داده شده است. مجموعه آیت‌های مکرر می‌توانند به گونه‌ای موثر هم در زمینه‌ی مکان و هم زمان، کاوش شوند. کارآمدی الگوریتم را می‌توان به طور نظری به وسیله‌ی ارتباط دادن تعداد مرزها به مفهوم جریان‌ها Farey از نظریه‌ی اعداد به دست آورد. از این ارتباط، ما نتیجه می‌گیریم که حتی در بدترین حالت، تعداد مرزها کندتر از طول جریان افزایش می‌یابد. آزمایش‌های روی مجموعه‌های داده‌ها، که در آن‌ها مجموعه آیت‌ها در الگوهای مختلفی رخ می‌دهند، در عمل، نسبت به زمان، نسبتاً پایا باقی می‌مانند. ما نشان دادیم که هیچ حاشیه‌ای نمی‌تواند از بهره‌برداری رابطه‌ی زیرمجموعه بین مجموعه آیت‌ها حاصل شود. به طور ساده، ذخیره‌سازی خلاصه‌های آیت منجر به مساعل فضا می‌شود. به اشتراک گذاری مرزها در میان خلاصه‌های ارتباط یافته توسط رابطه‌ی زیرمجموعه روی مجموعه آیت‌ها، غیر ممکن است زیرا در هر نقطه‌ی داده شده‌ای یک موقعیت ممکن است مرزی برای مجموعه آیت‌های  $A_1$ ،  $A_3$  باشد اما برای مجموعه آیت  $A_2$  نباشد، هنگامی که ما داشته باشیم  $A_1 \subseteq A_2 \subseteq A_3$ .

از آزمایشات با نمونه‌ی اولیه‌ی نرم‌افزار روی هر دوی مجموعه داده‌های جهان واقعی و مصنوعی، ما این سه نتیجه‌گیری را کسب کردیم. اولاً، روی میانگین، مرزهای بسیار کمی برای به دست آوردن ماکسیمم فرکانس همه‌ی مجموعه آیت‌ها در یک جریان  $V$  نیاز هستند. دوماً، جریان‌هایی که در کل دوره‌ی جریان، پراکنده و پشت سرهم هستند، از زمان‌بندی کنترل رسیدگی سست بهره می‌برند. سوماً، حتی جریان با معاملات بزرگ را می‌توان به طور موثر کاوش کرد اگر مقادیر مناسب برای طول پنجره‌ی مینیمال و آستانه‌ی فرکانس مینیمال، تنظیم شده باشند.



## توجه!

این فایل تنها قسمتی از ترجمه می‌باشد. برای تهیه مقاله ترجمه شده کامل با فرمت

ورد (قابل ویرایش) همراه با نسخه انگلیسی مقاله، [اینجا](#) کلیک نمایید.

برای جستجوی جدیدترین مقالات ترجمه شده، [اینجا](#) کلیک نمایید.